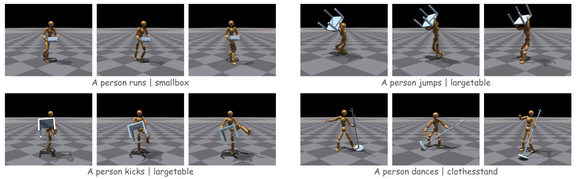
Dynamic Full-body Motion Agent with Object Interaction via Blending Pre-trained Modular Controllers

Sanghyeok Nam*, Byoungjun Kim*, Daehyung Park, Tae-Kyun Kim

KAIST

CVPR JUNE 3-7, 2026 DENVER COLORADO

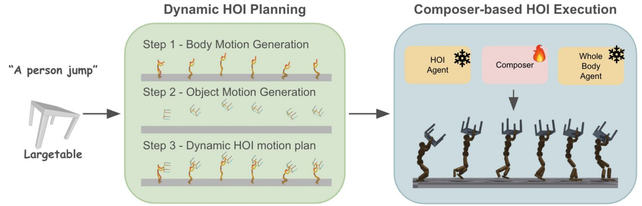
Introduction



run, jump, kick, dance while interacting with an object

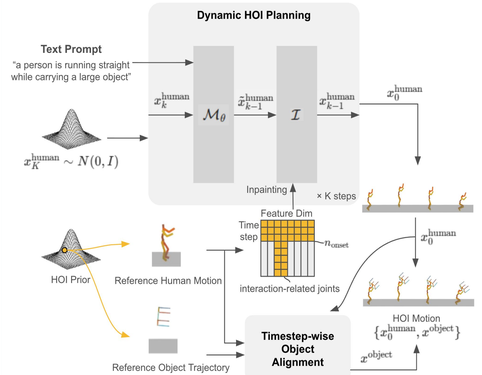
- Plan and execute physically plausible, **dynamic** and **contact-rich Human-Object Interaction (HOI)**
- Existing datasets and agents handle either HOI or dynamic motions, but **not both simultaneously**

Overview

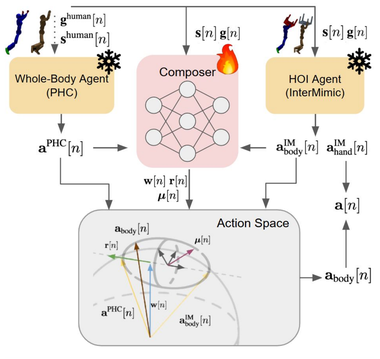


- **Dynamic HOI Planning**: Plan dynamic motions with consistent hand-object contact
- **Composer-based HOI Execution**: Efficiently learn dynamic HOI control by blending pre-trained experts

Methodology



Dynamic HOI Planning



$$\mathbf{a}_{\text{body}}[n] = \mathbf{a}^{\text{PHC}}[n] + (\mathbf{w}[n] + \mathbf{r}[n]) \odot \Delta\mathbf{a}_{\text{body}}[n] + \mathbf{U}\boldsymbol{\mu}[n]$$

$\mathbf{w}$: interpolation weights, $\mathbf{r}$: extrapolation weights, $\boldsymbol{\mu}$: exploration coeff., $\mathbf{U}$: PCA basis

Composer-based HOI Execution

Dynamic HOI Planning

- Inpaint HOI prior (FullBodyManip) into MDM during diffusion sampling
- Fix interaction-related joints after contact onset; text-aligned motion inpainting for the others
- Compute object trajectories from hand poses

Composer-based HOI Execution

- Composer predicts per-DoF inter-/extrapolation weights
- PCA-based low-dim subspaces for further exploration

Quantitative Results

Method	HOI	Physics	HOI QUALITY $C_{\%}\uparrow$	HOI QUALITY $C_{cons}\downarrow$	PHYSICAL PLAUSIBILITY $Pene_{obj}\downarrow$	PHYSICAL PLAUSIBILITY Skate ↓	PHYSICAL PLAUSIBILITY Float ↓	PHYSICAL PLAUSIBILITY $Jitter_{pos}\downarrow$	MOTION QUALITY R-Prec ↑	MOTION QUALITY Diversity ↑
MDM			–	–	–	**0.133**	29.3	9.43×10^4	**0.374**	**7.61**
HOI-Diff	✓		0.285	18.2	8.18	0.633	17.8	$\mathbf{1.56 \times 10^4}$	0.257	4.66
DAViD	✓		0.848	10.9	4.842	0.261	20.4	6.69×10^4	0.310	6.70
$Ours_P$	✓		**1.000**	**0.906**	4.196	0.217	30.1	5.93×10^4	0.332	5.56
$Ours_{P+E}$	✓	✓	0.999	2.95	**0.009**	0.786	**9.95**	4.53×10^4	0.316	3.97

Generation Quality

($Ours_P$: planning only / $Ours_{P+E}$: full planning-execution pipeline)

- $Ours_P$ achieves the **best HOI Quality** — $C_{\%}$ and C_{cons}
- $Ours_{P+E}$ achieves the **best physical plausibility** — $Pene_{obj}$ and Float, 2nd $Jitter_{pose}$

Method	$SR\uparrow$	T	$D\uparrow$	$E_{HOI}\downarrow$	$Jitter_{DoF}\downarrow$
PPO	0.227	75	0.859	59.098	0.424×10^3
PHC [26]	0.397	-	1.166	12.965	1.350×10^3
PHC_R	0.313	15	1.834	26.024	1.762×10^3
InterMimic [56]	0.376	-	2.407	21.379	0.436×10^3
$InterMimic_R$	0.310	15	2.324	34.675	$\mathbf{0.336 \times 10^3}$
$InterMimic_{FT}$	0.526	75	**7.37**	**8.635**	0.370×10^3
Ours	**0.591**	23	4.607	11.667	0.359×10^3

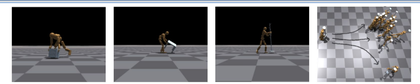
Blending	SR↑	D↑	$E_{HOI}\downarrow$
$Heuristic_{Hand}$	0.365	1.069	13.04
$Heuristic_{Arm}$	0.472	1.603	17.18
Hard MoE	0.383	2.748	20.48
Hard MoE (Joint)	0.383	2.535	26.45
$Ours_{MLP}$	0.571	3.034	18.03
$Ours_{MLP+PCA}$	**0.591**	**4.607**	**11.67**

Execution Performance

($Ours_{MLP}$: except PCA subspace exploration)

- Ours achieves the **highest success rate**
- Converges **3× faster** than fine-tuning InterMimic
- Ours blending method achieved best SR, D, and E_{HOI}

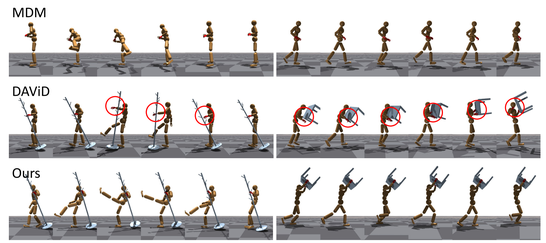
Scalability



(a) Push (b) Drag (c) Swing (d) Plan path

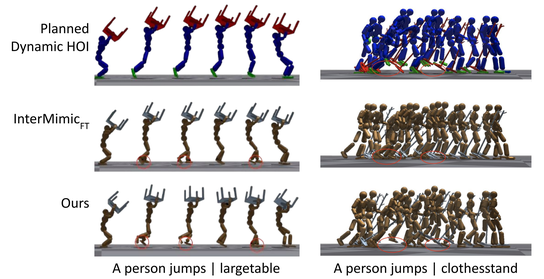
- Diverse object geometries (FullBodyManip)
- Diverse motion styles (push, drag, swing)
- Plug-and-play application — e.g., MDM → GMD enables path planning

Qualitative Results



Comparison of HOI Planning

- Ours accurately follows the text-aligned **dynamic motion** while maintaining **consistent hand-object contact**



Comparison of HOI Execution

- Ours successfully executes the planned dynamic HOI while maintaining **motion style** and **object interaction**
- $InterMimic_{FT}$ collapses to a local minimum — shuffling instead of two-foot takeoff

Contact

Sanghyeok Nam
KAIST, Republic of Korea
Email: sang990701@kaist.ac.kr
Website: https://www.linkedin.com/in/shnam99

References

1. N. Mahmood et al., "AMASS: Archive of Motion Capture as Surface Shapes," ICCV, 2019.
2. J. Li et al., "Object Motion Guided Human Motion Synthesis," ACM TOG, 2023.
3. G. Tevet et al., "Human Motion Diffusion Model," ICLR, 2023.
4. X. Peng et al., "HOI-Diff: Text-Driven Synthesis of 3D Human-Object Interactions Using Diffusion Models," CVPR Workshop on HuMoGen, 2025.
5. H. Kim et al., "DAVID: Modeling Dynamic Affordance of 3D Objects Using Pre-trained Video Diffusion Models," ICCV, 2025.
6. Z. Luo et al., "Perpetual Humanoid Control for Real-time Simulated Avatars," ICCV, 2023.
7. S. Xu et al., "InterMimic: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions," CVPR, 2025.
8. K. Karunratanakul et al., "Guided Motion Diffusion for Controllable Human Motion Synthesis," ICCV, 2023.